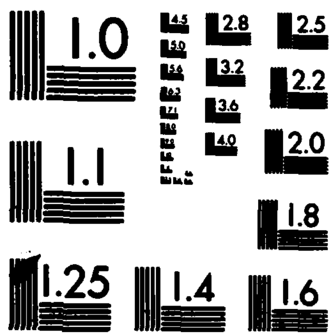
AD-A138 351 BAYES LEAST SQUARES LINEAR REGRESSION IS ASYMPOTICALLY FULL BAYES: ESTIMA..(U) OREGON STATE UNIV CORVALLIS DEPT OF STATISTICS H D BRUNK ET AL. JAN 84 1/1

UNCLASSIFIED OSU-ONR-TR-84-1 N00014-81-K-0814 F/G 12/1 NL

END
FILMED




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA138351

OSU - ONR TR 84-1

BAYES LEAST SQUARES LINEAR REGRESSION IS ASYMPOTICALLY FULL BAYES:

ESTIMATION OF SPECTRAL DENSITIES

H.D. Brunk
Department of Statistics
Oregon State University
Corvallis, Oregon 97331

Submitted for publication to:

J. American Statistical Association

Prepared for:

The Office of Naval Research

under

Contract No. N00014-81-K0814

R.R. Mohler, Principal Investigator



January 1984

# BAYES LEAST SQUARES LINEAR REGRESSION IS ASYMPTOTICALLY FULL BAYES: ESTIMATION OF SPECTRAL DENSITIES

## 1. INTRODUCTION

Bayes least squares linear (BLSL) estimators were introduced by Whittle (1957, 1958) and described explicitly and further developed by Hartigan (1969). The method was applied to estimation of coefficients of orthogonal expansions of regression functions in another work ~~(Brunk, 1980)~~. In this ~~the~~ present paper it is noted ~~we observe~~ that when many observations are available we can expect the BLSL method can be expected to yield substantially the same results as a full Bayesian treatment; and ~~we illustrate~~ the method is illustrated in the context of estimation of spectral densities. In that context, the estimators suggested will appear rather ordinary. But they are not completely ad hoc: each comes with an interpretation. And, when large samples are available, the posterior distribution of the estimator at a fixed frequency is (approximately) normal, with easily calculated standard deviation.

## 2. ORTHOGONAL EXPANSIONS AND THE BLSL METHOD

In order to be more explicit, we recall the description of the estimators given in (Brunk, 1980). Since the size of the data set is relevant here, we allow the number of observations to appear in the notation. Thus for each integer $n$ we have a set $\{x_{no}, x_{n1}, \ldots, x_{nn}\}$ of values of an "explanatory variable" in a space $X$ of possible values. The regression function $R_n$ is defined on $X$ and is assumed to have a finite expansion in terms of specified functions $r_{no}$, $R_{no}$ and $\{\phi_{nr}, r = 0, 1, \ldots, n\}$:

$$R_n(x) = R_{no}(x) + r_{no}(x) \sum_{r=0}^{n} \beta_{nr}\, \phi_{nr}(x), \quad x \in X. \tag{2.1}$$

The observations or responses $(Y_{no}, Y_{n1}, \ldots, Y_{nn})$ are assumed independent, and $Y_{nj}$ is assumed to have mean

$$E(Y_{nj}) = R_n(x_{nj}) \tag{2.2}$$

and variance

$$\mathrm{Var}(Y_{nj}) = 1/\lambda_n \pi_{nj} \quad , \quad j = 0, 1, \ldots, n, \tag{2.3}$$

where $\pi_{no}, \ldots, \pi_{nn}$ are known. We shall assume also that $\lambda_n$ is known, though in practice $\lambda_n$ may be estimated from the data.

(One can state less restrictive assumptions that suffice. In what follows, as elsewhere, a tilde underline indicates a random entity, and " := " is used between two expressions when the left is defined by the right. These assumptions are that the linear expectation of $Y_{nj}$ be $R_n(x_{nj})$:

$$LE(Y_{nj} | \underset{\sim}{\beta}_n = \beta_n) = R_n(x_{nj}), \quad j = 0, 1, \ldots, n,$$

where $\underset{\sim}{\beta}_n := (\underset{\sim}{\beta}_{no}, \ldots, \underset{\sim}{\beta}_{nn})'$ ; and the linear covariance matrix of $Y_n := (Y_{no}, \ldots, Y_{nn})'$ given $\underset{\sim}{\beta}_n = \beta_n$ have entries

$$E([Y_{ni} - R_n(x_{ni})][Y_{nj} - R_n(x_{nj})] | \underset{\sim}{\beta}_n = \beta_n) = \delta_{ij} / \lambda_n \pi_{nj},$$

$$i, j = 0, 1, \ldots, n;$$

here $\delta_{ij}$ is the Kronecker delta: $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$.)

The functions $\{\phi_{nr}, r = 0, 1, \ldots, n\}$ are assumed selected so as to be orthonormal with respect to the design of the experiment:

$$\sum_{j=0}^{n} \pi_{nj} r_{no}^{2}(x_{nj})\ \phi_{nr}(x_{nj})\ \phi_{ns}(x_{nj}) = \delta_{rs}\ . \tag{2.4}$$

The function $R_{no}$ is thought of as a prior mean, so that one sets

$$E(\underset{\sim}{\beta}_{nr}) = 0 \quad , \quad r = 0, 1, \ldots, n\ . \tag{2.5}$$

It is argued in (Brunk, 1980) that since each coefficient $\beta_{nr}$ has an interpretation independent of $\beta_{ns}$ for $s \neq r$, it may often be reasonable to assign $\underset{\sim}{\beta}_n := (\underset{\sim}{\beta}_{ns}, \underset{\sim}{\beta}_{ni}, \ldots, \underset{\sim}{\beta}_{nn})'$ a joint prior distribution according to which the components $\underset{\sim}{\beta}_{no}, \ldots, \underset{\sim}{\beta}_{nn}$ are independent; and we set

$$\tau_{nr} := 1/\mathrm{Var}(\underset{\sim}{\beta}_{nr})\ ,\ r = 0, 1, \ldots, n\ . \tag{2.6}$$

Set

$$\begin{aligned} U_{nr} &:= \sum_{j=0}^{n} \pi_{nj} r_{no}(x_{nj})\ \phi_{nr}(x_{nj})[Y_{nj} - R_{no}(x_{nj})] \\ &= \sum_{j=0}^{n} c_{nrj} W_{nj}\ , \quad r = 0, 1, \ldots, n\ , \end{aligned} \tag{2.7}$$

where

$$c_{nrj} := \sqrt{\pi_{nj}}\ r_{no}(x_{nj})\ \phi_{nr}(x_{nj}) \tag{2.8}$$

and

$$W_{nj} := \sqrt{\pi_{nj}}\,[Y_{nj} - R_{no}(x_{rj})]\,, \quad r, j = 0, \ldots, n\,. \tag{2.9}$$

That is, the random vector

$$U_n := (U_{no}, U_{nr}, \ldots U_{nn})'$$

is obtained by applying the orthogonal transformation $C_n$ to the vector $W_n := (W_{no}, \ldots, W_{nn})'$ , where

$$(C_n)_{rj} := c_{nrj}\,, \quad r = 0, 1, \ldots, n\,, \quad j = 0, 1, \ldots, n\,.$$

It follows from (2.4) that

$$E(U_{nr} | \underset{\sim}{\beta}_n = \beta_n) = \beta_{nr} \tag{2.10}$$

and that

$$[\mathrm{cov}(U_{nr}, U_{ns}) | \underset{\sim}{\beta}_n = \beta_n] = \delta_{rs} / \lambda_n\,. \tag{2.11}$$

Then the linear expectation of $\underset{\sim}{\beta}_n$ given $U_{no}, \ldots, U_{nn}$ is given by

$$\hat{\underset{\sim}{\beta}}_{nr} = \lambda_n U_{nr} / (\lambda_n + \tau_{nr})\,, \quad r = 0, 1, \ldots, n \tag{2.12}$$

and the linear variances and covariances are given by

$$E(\underset{\sim}{\beta}_{nr} - \hat{\underset{\sim}{\beta}}_{nr})^2 = 1/(\lambda_n + \tau_{nr})\,, \qquad r = 0, 1, \ldots, n\,, \tag{2.13}$$

$$E(\underset{\sim}{\beta}_{nr} - \hat{\underset{\sim}{\beta}}_{nr})(\underset{\sim}{\beta}_{ns} - \hat{\underset{\sim}{\beta}}_{ns}) = 0 \quad \text{for} \quad r \neq s\,. \tag{2.14}$$

For fixed $x$ , the linear expectation of $R_n(x)$ is

$$\hat{R}_n(x) = R_{no}(x) + r_{no}(x) \sum_{r=0}^{n} \hat{\beta}_{nr} \phi_{nr}(x) \tag{2.15}$$

and its linear variance is

$$E([R_n(x) - \hat{R}_n(x)]^2) = r_{no}^2(x) \sum_{r=0}^{n} \phi_{nr}^2(x)/(\lambda_n + \tau_{nr}) . \tag{2.16}$$

Note that the method does not, in general, provide posterior covariances.

It will be useful to note that when $R_{no} = 0$ and $\tau_{nj} = 0$ for $j = 0, 1, \ldots, n$, the formula for $\hat{R}_n(x)$ provides the ordinary least squares regression of $Y$ on $x$ with weights $\pi_{nj}$, $j = 0, 1, \ldots, n$. Indeed, when $k$ is fixed, $k <= n$, the orthogonality properties of the functions $\{\phi_{nr} : r = 0, 1, \ldots, n\}$ lead to $\sum_{r=0}^{k} \hat{a}_{nr} \phi_{nr}$ as ordinary least squares estimator of $R_n$, where

$$\hat{a}_{nr} := \sum_{j=0}^{n} \pi_{nj} \phi_{nr}(x_{nj}) Y_{nj} , \quad r = 0, 1, \ldots, k .$$

Note that $\hat{a}_{nr} = U_{nr}$ if $R_{no} = 0$ and $\hat{a}_{nr} = \hat{\beta}_{nr}$ if also $\tau_{nr} = 0$, $r = 0, 1, \ldots, k$.

Of course, the BLSL estimators may be considered from a conventional point of view. That is, one may choose, as is customary when estimating a regression function $R_n$, some family of functions considered adequate for representing it. One may then orthogonalize these functions with respect to the design, to obtain functions $\{\phi_{nr}, r = 0, 1, \ldots, n\}$; and then specify a function $R_{no}$ and "precisions" $\{\tau_{nr}, r = 0, 1, 2, \ldots, n\}$, thus finally obtaining an estimator of the regression function $R_n$. But such an estimator is not completely ad hoc; it comes with an interpretation. One realizes that one is considering the same estimator that another investigator would use who was applying the Bayes least squares linear method, specifying $R_{no}$ as a prior mean, and the $\{\tau_{nr}, r = 0, 1, \ldots, n\}$ as precisions of the coefficients in the expansion of $R_n$. And one may like to bear that in mind when specifying $R_{no}$ and the precisions.

In principle, an investigator who is well acquainted with the functions $\{\phi_{nr}\}$ to be used, and who also has a clear and definite opinion as to the probable shape of the regression function to be estimated, can specify, more or less uniquely, a prior mean and prior precisions. But in practice, there may be a rather wide variety of specifications that all seem reasonable. One may then examine a number of estimates arising from a range of "reasonable" priors. As to the specification of the prior mean, $R_{no}$, a heuristic argument is given in (Brunk, 1981) that it is often reasonable to fit the data--roughly--by a member of a family of smooth functions depending on only one or two parameters. That leaves still the precisions, $\tau_{nr}$, $r = 0, 1, \ldots, n$. Two suggestions come from thinking of them as reciprocals of prior variances of the parameters $\underset{\sim}{\beta}_{nr}$, $r = 0, 1, 2, \ldots, n$.

(i) If the functions $\phi_{nr}$ oscillate more and more rapidly with increasing $r$, one can express an opinion that the estimate $\hat{R}_n$ is "smooth" by specifying large values of $\tau_{nr}$ when $r$ is large.

(ii) The precisions should be specified independently of the data.

## 3. APPROXIMATE NORMALITY

Now let us briefly imagine that the random variables $U_{no}, \ldots, U_{nn}$ were observed (rather than $Y_{no}, \ldots, Y_{nn}$), that they were independent according to their joint distribution given $\underset{\sim}{\beta}_n$, and that

$$U_{nr} | \underset{\sim}{\beta}_n = \beta_n \sim N(\beta_{nr}, \ 1/\lambda_n) . \tag{3.1}$$

Suppose also that $\underset{\sim}{\beta}_n$ is given a joint prior distribution according to which its components $\underset{\sim}{\beta}_{no}, \ldots, \underset{\sim}{\beta}_{nn}$ are independent, and

$$\underset{\sim}{\beta}_{nr} \sim N(0, \ 1/\tau_{nr}), \ r = 0, 1, \ldots, n . \tag{3.2}$$

Then these components are also independent and normally distributed according to their posterior distribution, with

$$E(\underset{\sim}{\beta}_{nr} | U_{no} = u_{no}, \ldots, U_{nn} = u_{nn}) = \lambda_n u_{nr}/(\lambda_n + \tau_{nr}) , \tag{3.3}$$

(cf. (2.12) ) and

$$\operatorname{Var}(\underset{\sim}{\beta}_{nr} | U_{no} = u_{no}, \ldots, U_{nn} = u_{nn}) = 1/(\lambda_n + \tau_{nr}) , \quad r = 0, 1, \ldots, n \tag{3.4}$$

(cf (2.13), (2.14) ) .

We are interested particularly in contexts in which one expresses an opinion as to the "smoothness" of $R_n$ by assigning large precisions $\tau_{nr}$ to the coefficients of rapidly varying functions $\phi_{nr}$ in the expansion of $R_n$. Then, typically, there is a positive integer $m$ such that the posterior mean and variance of $\underset{\sim}{\beta}_{nr}$ are so near zero for $r > m$ that corresponding terms in the expansion can be neglected; and this is so also for the posterior linear expectation and for $E(\hat{\underset{\sim}{\beta}}_{nr} - \hat{\underset{\sim}{\beta}}_{nr})^2$ in the BLSL method.

When the observations $Y_{no}, \ldots, Y_{nn}$ are <u>not</u> jointly normal, but $n$ is large, we shall argue that often $U_{no}, \ldots, U_{nn}$ will have approximately a joint normal distribution (cf. Theorem 4.1, to follow). The data $U_{no}, \ldots, U_{nn}$ are fully equivalent to the original data $Y_{no}, \ldots, Y_{nn}$ : each may be obtained from the other by an orthogonal linear transformation. And for some positive integer $m$, $U_{n,m+1}, \ldots, U_{nn}$ may safely be ignored, so that if $\underset{\sim}{\beta}_{no}, \ldots, \underset{\sim}{\beta}_{nn}$ are given the multinormal prior distribution described above, then according to their joint posterior distribution they are (approximately) jointly normally distributed with posterior means given by (2.12) and (3.3) and posterior variances by (2.13) and (3.4). A theorem (Theorem 4.1) that suggests this approximation is given in Section 4, and its proof in the appendix.

## 4. ASYMPTOTIC NORMALITY

We consider triangular arrays $\mathscr{V}$ : $V_{no}, V_{n1}, \ldots, V_{n,kn}$, $n = 1, 2, \ldots,$ where $k_n \to \infty$ as $n \to \infty$, and where $E(V_{nj}) = 0$, $\mathrm{Var}(V_{nj}) = 1$, $j = 0,1, \ldots, k_n$. We shall argue that this array is asymptotically normal, given $\underset{\sim}{\beta}_n = \beta_n$, when

$$V_{nr} := [U_{nr} - E(U_{nr})]/[\mathrm{Var}(U_{nr})]^{\frac{1}{2}}, \quad r = 0,1, \ldots, k_n \tag{4.1}$$

provided that $k_n$ does not grow too fast. We use Mallows's (1972) definition of asymptotic normality: the array $\mathscr{V}$ is jointly asymptotically normal (j.a.n.) if for every array, $\mathcal{A}$, of reals $a_{no}$, $a_{nj}$, ..., $a_{n,kn}$, $n = 0,1,\ldots$, such that $\sum_{r=0}^{k_n} a_{nr}^2 = 1$, the random variable $\sum_{r=o}^{k_n} a_{nr} V_{nr}$ converges in distribution to the standard normal distribution. (Mallows (1972) observes that this implies that for each $d$, $(V_{no}, \ldots, V_{nd})$ converges in distribution to the standard d-dimensional normal distribution.)

<u>Theorem</u> 4.1. Let there be a positive number $M$ such that

$$(\lambda_n \pi_{nj})^{3/2} E|Y_{nj} - \mu_{nj}|^3 <= M, \tag{4.2}$$

where

$$\mu_{nj} := E(Y_{nj}) = R_n(x_{nj}), \quad j = 0,1,\ldots,n, \quad n = 1,2,\ldots. \tag{4.3}$$

And suppose that

$$\max\{\sqrt{\pi_{nj}}\, |r_{no}(x_{nj})| \sum_{r=0}^{k_n} |\phi_{nr}(x_{nj})| : j = 0,1,\ldots,n\} \to 0 \tag{4.4}$$

$$\text{as} \quad n \to \infty .$$

Then the array $\mathcal{V}$ is j.a.n.

The proof, given in the appendix, consists of showing in a straightforward way that the characteristic function of $\sum_{r=0}^{k_n} a_{nr} V_{nr}$, evaluated at a real number $t$, converges to $\exp(-t^2/2)$ as $n \to \infty$. This theorem presents an instance of a phenomenon studied by Mallows (1969). Note that the random variables $U_{nr}$ are obtained via an orthogonal transformation from the random variables $W_{nj}$; and that while these are independent, they need not be normally distributed. Although the inverse orthogonal transformation would recover the original non-normal random variables, the random variables $U_{nr}$ are nevertheless j.a.n. Mallows (1969) proves a theorem with a somewhat stronger conclusion, described in terms of the integrated squared difference between the standard normal distribution function and the distribution function it approximates. As stated, Mallows's theorem requires independent, identically distributed random variables. While the method of proof appears to allow a relaxation of that requirement, it does seem to require that the distributions not be too nearly of lattice type.)

## 5. BLSL ESTIMATION OF SPECTRAL DENSITIES

We consider the problem of estimating the spectral density of a stationary, purely nondeterministic time series

$$X_t = \sum_{s=0}^{\infty} \alpha_s \; \varepsilon_{t-s}$$

(not necessarily Gaussian), where the random variables $\underset{\sim}{\varepsilon}_t$, $t = \ldots, -1, 0, 1, \ldots$ are independent with zero means and common variance. The coefficients $\alpha_s$, $s = 0, 1, \ldots,$ are unknown, but we assume that $\sum_{t=0}^{\infty} |\alpha_t| < \infty$ and that

$$\sup_t E[\, \underset{\sim}{\varepsilon}_t^2 \, 1_{\{|\underset{\sim}{\varepsilon}_t| > c\}}] \longrightarrow 0 \quad \text{as} \quad c \longrightarrow \infty$$

(cf. Anderson (1971), page 482). The spectral density, to be estimated, is

$$f(\omega) = \sum_{t=-\infty}^{\infty} r(t)\exp(2\pi i\omega t)\ , -1/2 <= \omega <= 1/2\ ,$$

where

$$r(t) := E(X_s X_{s+t})$$

is the covariance function. We follow Wahba (1980) in formulating the problem as an ordinary regression problem, with values of the log periodogram (cepstrum) as data. Let $X_1, X_2, \ldots, X_{2n}$ be observed. We have

$$I_n(\omega) := (1/2n) \left| \sum_{i=1}^{2n} X_t \exp(2\pi i\omega t) \right|^2 , \quad -1/2 <= \omega <= 1/2 ,$$

$$I_{nj} := I_n(j/2n) \quad (= I_n(-j/2n))$$

$$= f(j/2n) T_{nj} ,$$

where

$$T_{nj} := I_{nj} / f(j/2n) , \quad j = 0, 1, 2, \ldots, n .$$

Asymptotically, as $n \to \infty$, the random variables $\{T_{nj}, j = 0,1,\ldots,n\}$ are independent, with $T_{no}$ and $T_{nn}$ distributed as $\chi^2(1)$ and $2T_{nj} \sim \chi^2(2)$ for $j = 1,2,\ldots,n-1$ (Anderson, 1971, pp. 484-485). We set

$$Y_{nj} := \log I_{nj} + C_j , \quad j = 0,1,\ldots,n ,$$

where $C_o = C_n := (\ln 2 + \gamma)$, $C_j := \gamma$, $j = 1,2, \ldots, n-1$, and where $\gamma$ is the Euler-Mascheroni constant, approximately 0.57721. Then the random variables $\{Y_{nj}, j = 0,1, \ldots, n\}$ are asymptotically independent, with $\{Y_{nj}, j = 1,2, \ldots, n-1\}$ asumptotically identically distributed.

We set

$$R_n(x) := \log f(x/2) \quad \text{for} \quad 0 <= x <= 1 .$$

According to the asymptotic distribution, According to the asymptotic distribution,

$$E(Y_{nj}) = R_n(j/n) = R_n(x_{nj}) ,$$

where $x_{nj} := j/n$ for $j = 0, 1, 2, \dots, n$, while

$$\mathrm{Var}(Y_{nj}) = \pi^2/6 , \quad j = 1, 2, \dots , n-1 ,$$

$$\mathrm{Var}(Y_{nj}) = \pi^2/2 \quad \text{for} \quad j = 0, n .$$

We shall conduct the analysis and carry out the computations as if $\mathrm{Var}(Y_{nj}) = \pi^2/3$ for $j = 0, n$, since in any case the influence of these two terms is negligible for large $n$. In the notation of Section 2, then, we take

$$\left.\begin{array}{l} \pi_{nj} := \pi_{nn} := 1/2 , \quad \pi_{nj} := 1 \quad \text{for} \quad j = 1,2, \dots , n-1 , \\ \text{and} \\ \lambda_n := 6/\pi^2 , \quad n = 1,2, \dots . \end{array}\right\} \tag{5.1}$$

The functions $\phi_{nr}$ are given by

$$\phi_{no}(x) := 1/\sqrt{n} ,$$

$$\phi_{nr}(x) := (2/n)^{\frac{1}{2}} \cos \pi r x , \quad r = 1,2, \dots , n , \quad x \in [0,1] .$$

The function $r_{no}$ used is identically 1.

It is most convenient to take as $R_o$ a linear combination of the functions $\{\phi_{nr} : r = 0, 1, \dots , n\}$. As we noted earlier, a heuristic argument is given in (Brunk, 1981, page 117) that it is reasonable for the investigator to select as prior mean $R_o$ a regression function that is consistent with both the data and his opinion as to its shape. In particular, the investigator may simply use ordinary least squares with weights $\pi_{no}, \ldots, \pi_{nn}$ to choose coefficients $a_o, a_1, \ldots, a_k$ for $k = 1$ or $2$ or $3$ in fitting the function $\sum_{r=0}^{k} a_r \phi_{nr}$ to the data. This would yield

$$R_o(x) = \sum_{r=0}^{k} a_r \phi_{nr}(x)$$

where

$$a_r := \sum_{j=0}^{n} \pi_{nj} \phi_{nr}(x_{nj}) Y_{nj} .$$

And then (cf. the end of Section 2) if $\tau_0 = \tau_1 = \cdots = \tau_k = 0$, we have

$$\hat{R}_n(x) = \sum_{r=0}^{n} \hat{\beta}'_{nr} \phi_{nr}(x) , \text{ where}$$

$$\hat{\beta}'_{nr} := \lambda_n U'_{nr} / (\lambda_n + \tau_{nr}) , \quad r = 0, 1, \ldots, n ,$$

and

$$U'_{nr} := \sum_{j=0}^{n} \pi_{nj} \phi_{nr}(x_{nj}) Y_{nj} , \; r = 0, 1, \ldots, n .$$

In other terms, if the investigator chooses to consider a specification of prior mean and precisions that has $R_0$ as the ordinary least squares estimator of $R_n$ as a linear function of $\phi_{no}, \phi_{n1}, \ldots, \phi_{nk}$, and has prior precisions $\tau_{no} = \tau_{n1} = \ldots \tau_{nk} = 0$, then $\hat{R}_n$ is precisely what it would be if he took $R_o \equiv 0$ (in which case $U_{nr}$ would become $U'_{nr}$).

The functions $\phi_{nr}$ described above depend--in a simple way--on $n$. It is more convenient when considering specification of precisions to use functions independent of $n$: $\phi^*_n := (n/2)^{\frac{1}{2}} \phi_{nr}$, so that

$$\phi_o^* := \sqrt{1/2}, \quad \phi_r^*(x) := \cos \pi r x, \quad r > 0. \tag{5.2}$$

We set

$$\beta_{nr}^* := (2/n)^{\frac{1}{2}}\beta_{nr}, \quad r = 0, 1, \ldots, n.$$

Then--n being fixed--we are assuming that $R_n$ has an expansion

$$R_n(x) = R_{no}(x) + \sum_{r=0}^{n} \beta_{nr}^* \phi_t^*(x), \quad x \in [0,1].$$

While formally the assumed expansion of $R_n$ depends on $n$, we consider that terms of large index are negligible; and each term is in fact independent of $n$, so that we write

$$R(x) = R_o(x) + \sum_{r=0}^{n} \beta_r^* \phi_r^*(x), \quad x \in [0,1]. \tag{5.3}$$

The coefficients $\{\beta_r^*, r = 0,1, \ldots, n\}$, modeled as random variables, have means

$$E(\underset{\sim}{\beta}_r^*) = 0$$

and precisions

$$\tau_r^* := 1/\mathrm{Var}(\underset{\sim}{\beta}_r^*).$$

In the present context and notation, Equation (2.12) becomes

$$\hat{\beta}_{nt}^* = U_r^* / (n/2 + \pi^2 \tau_r^* / 6), \tag{5.4}$$

where

$$U_r^* := \sum_{j,o}^{n} \pi_{nj} \phi_r^* (j/n) [Y_{nj} - R_o(j/n)],$$
$$r = 0,1, \ldots, n. \tag{5.5}$$

We note that $E(2U_r^* / n | \underset{\sim}{\beta}^*) = \underset{\sim}{\beta}_r^*$, and that $2U_r^*/n$ is the ordinary least squares estimate of $\beta_r^*$ with weights $\pi_{ni}$, $i = 0,1, \ldots, n$. For fixed $x \in [0,1]$, the posterior linear expectation of $R(x)$ is

$$\hat{R}_n(x) = R_o(x) + \sum_{r=0}^{n} \underset{\sim}{\beta}_{nr}^* \phi_r^*(x) \tag{5.6}$$

and its linear variance is

$$E[R(x) - \hat{R}_n(x)]^2 = \sum_{r=0}^{n} [\phi_r^*(x)]^2 / (3n/\pi^2 + \tau_r^*) . \tag{5.7}$$

In view of Theorem 4.1, when $n$ is large we expect the posterior distribution of $R(x)$ to be approximately normal with mean $\hat{R}_n(x)$ given by (5.6) and variance given by (5.7).

## 6. EXAMPLES

As a first example, we have used the example used by Wahba (1980):

$$X_t = \sum_{k=1}^{3} \gamma_k X_{t-k} + \underset{\sim}{\varepsilon}_t ,$$

where $\gamma_1 = 1.4256$, $\gamma_2 = -0.7344$, $\gamma_3 = 0.1296$, and where the random variables $\underset{\sim}{\varepsilon}_t$, $t = \ldots, -2, -1, 0, 1, 2, \ldots$ are independent, each having the standard normal distribution. The simulation was carried out starting with $X_{-30} = 0$ and then discarding $X_{-30}, X_{-29}, \ldots, X_0$. (These observed $X_t$ are not to be confused with $x_{nj} := j/n$ in the formulas in Section 5.) One set of 256 points was obtained in this way $(n = 128)$, as also a larger set of 1024 $(n = 512)$ containing the first.

The function $R_o$ used is defined by:

$$R_o(x) := 0.1 + 2.9\cos(\pi x) + 0.5\cos(2\pi x),\ 0 \leq x \leq 1 .$$

The accompanying Tables 1 and 2, and Figures 1 through 14, relate to the following four specifications of precisions $\tau_r^*$, $r = 0, 1, \ldots, n$:

A: $\tau_r^* := (0.21\ r)^8$ ,

B: $\tau_r^* := (0.00024)(6.4)^r$ ,

C: $\tau_r^* := 0.004(4)^r$ ,

and

D: $\tau_r^* := 0.1(3)^r$ .

Table 1 indicates the "damping" effect of each specification of precisions; that is, the entry in the table is $1/(1+\pi^2\tau_r^*/3n)$ , the factor by which the ordinary least squares estimate , $2U_r^*/n$ , is multiplied to obtain $\hat{\beta}_r^*$ , when $n = 128$ (256 observations). The entries in Table 2 are for $n = 512$ (1024 observations). Each specification of $\{\tau_r^*,\ r = 0, 1, \ldots ,\}$ leads to a "window estimator" that could be considered from a conventional point of view. Any or all of these specifications might appear reasonable to an investigator. The precisions specified under D increase most rapidly, and might be expected to lead to the smoothest estimates of $R$ . Initially, the precisions A increase somewhat more rapidly than those of B , though eventually those of B increase much more rapidly. In fact, those of B were deliberately selected (by regression of log $\tau_r^*$ on $r$ , from A) so as to be near those of A for $r <= 10$ .

Table 1

Multipliers for n = 128

| r | A | B | C | D |
|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 1.000 | 0.997 |
| 1 | 1.000 | 1.000 | 1.000 | 0.992 |
| 2 | 1.000 | 1.000 | 0.998 | 0.977 |
| 3 | 0.999 | 1.000 | 0.993 | 0.935 |
| 4 | 0.994 | 0.999 | 0.974 | 0.828 |
| 5 | 0.963 | 0.993 | 0.905 | 0.616 |
| 6 | 0.860 | 0.959 | 0.704 | 0.348 |
| 7 | 0.641 | 0.787 | 0.373 | 0.151 |
| 8 | 0.380 | 0.365 | 0.129 | 0.056 |
| 9 | 0.193 | 0.083 | 0.036 | 0.019 |
| 10 | 0.093 | 0.014 | 0.009 | 0.007 |
| 11 | 0.046 | 0.002 | 0.002 | 0.002 |
| 12 | 0.023 | 0.000 | 0.001 | 0.001 |
| 13 | 0.012 | 0.000 | 0.000 | 0.000 |
| 14 | 0.007 | 0.000 | 0.000 | 0.000 |
| 15 | 0.004 | 0.000 | 0.000 | 0.000 |
| 16 | 0.002 | 0.000 | 0.000 | 0.000 |
| 17 | 0.001 | 0.000 | 0.000 | 0.000 |
| 18 | 0.001 | 0.000 | 0.000 | 0.000 |
| 19 | 0.001 | 0.000 | 0.000 | 0.000 |
| 20 | 0.000 | 0.000 | 0.000 | 0.000 |

Table 2

Multipliers for n = 512

| r | A | B | C | D |
|---|---|---|---|---|
| 0 | 1.000 | 1.000 | 1.000 | 0.999 |
| 1 | 1.000 | 1.000 | 1.000 | 0.998 |
| 2 | 1.000 | 1.000 | 1.000 | 0.994 |
| 3 | 1.000 | 1.000 | 0.998 | 0.983 |
| 4 | 0.998 | 1.000 | 0.993 | 0.951 |
| 5 | 0.991 | 0.998 | 0.974 | 0.865 |
| 6 | 0.961 | 0.990 | 0.905 | 0.681 |
| 7 | 0.877 | 0.936 | 0.704 | 0.416 |
| 8 | 0.710 | 0.697 | 0.373 | 0.192 |
| 9 | 0.489 | 0.265 | 0.129 | 0.073 |
| 10 | 0.292 | 0.053 | 0.036 | 0.026 |
| 11 | 0.161 | 0.009 | 0.009 | 0.009 |
| 12 | 0.087 | 0.001 | 0.002 | 0.003 |
| 13 | 0.048 | 0.000 | 0.001 | 0.001 |
| 14 | 0.027 | | | |
| 15 | 0.016 | | | |
| 16 | 0.009 | | | |
| 17 | 0.006 | | | |
| 18 | 0.004 | | | |
| 19 | 0.002 | | | |
| 20 | 0.002 | | | |

Figure 1 shows both the true $R(x) := \log f(x/2)$ and the estimate $\hat{R}$ obtained from the 256 observations $(n = 128)$ using precisions A. Figures 2, 3, and 4 show estimates obtained through precisions B, C, and D respectively. Figure 5 shows first and fourth estimates, together with the true $R$. Figure 6 provides the graphs for the first prior, but based on the 1024 observations $(n = 512)$, and Figure 7 is for the fourth prior. Figure 8 shows first and fourth together.

Figure 9 shows the spectral density estimate and the true spectral density, graphed against twice the frequency, for the first prior (A), and 256 observations. The curves lying above and below the graph of the estimate indicate the precision of the estimate in the following way. For fixed $x$, the asymptotic theory leads us to expect $\hat{R}(x)$ to be approximately normally distributed according to its posterior distribution. Its (approximate) posterior variance is given by (5.7). Letting $\sigma(x)$ denote the square root of this posterior variance, the upper and lower graphs are graphs of $\exp[\hat{R}(x) \pm \sigma(x)]$. Figure 10 gives the same information for the fourth prior (D), and Figure 11 shows first and fourth estimates together. Figures 12 and 13 compare with Figures 9 and 10, but for the case of 1024 observations. Spectral density estimates for first and fourth priors are shown together in Figure 14, for the case of 1024 observations.

For a second example we have used underwater ambient noise data kindly furnished by the Naval Undersea Warfare Experiment Station at Keyport, Washington. A sample of 1024 observations

($n = 512$) was taken, with effective sampling frequency 80 khz. The procedure described in Section 5 was followed, and precisions $\tau_r^* := 0.1(3^*)$, $r = 0, 1, \ldots, 512$ were used. Figure 15 shows the estimated log spectral density, together with graphs obtained by adding and by subtracting the square root of the expected squared error; i.e., (approximately) one standard deviation of the (approximately) normal posterior distribution of the estimate for fixed frequency. Figure 16 shows the corresponding graphs for the spectral density; each ordinate in Figure 15 is just the natural logarithm of the corresponding ordinate in Figure 16.

## 7. Acknowledgments

The author gratefully acknowledges the support of the Office of Naval Research through Contract N00014-81-K-0814 with Ronald Mohler as principal investigator, and the National Science Foundation through Grant MCS 80 02907-01; and the cooperation of the Naval Underwater Warfare Experiment Station at Keyport, Washington. He wishes to thank Richard Bucolo and Roy Rathja for their assistance; and most particularly Ronald Stillinger for most effective use of his programming expertise.

# REFERENCES


Anderson, T. W. (1971). The Statistical Analysis of Time Series. John Wiley and Sons, Inc.

Brunk, H. D. (1980). Bayesian least squares estimates of univariate regression functions. Comm. Statist. -Theor. Meth., A9(11), 1101-1136.

__________ (1981). Estimation of stimulus-response curves by Bayesian least squares. Psychometrika 46, 115-128.

Chung, K. L. (1968). A Course in Probability Theory. Harcourt, Brace, and World.

Hartigan, J. A. (1969). Linear Bayes Methods. J. Royal Statist. Soc. B, 31, 446-454.

Mallows, C. L. (1969). Joint normality induced by orthogonal transformations. Bell Telephone Laboratories Memorandum.

__________ (1972). Linear processes are nearly Gaussian. J. Appl. Prob. 4, 313-329.

Wahba, G. (1980). Automatic smoothing of the log periodogram. J. Amer. Statist. Ass'n. 75, 122-132.

Whittle, P. (1957). Curve and periodogram smoothing. J. Royal Statist. Soc. B, 19, 38-46.

__________ (1958). On the smoothing of probability density functions. J. Royal Statist. Soc. B, 20, 334-343.

## APPENDIX

Proof of Theorem 4.1.

Set

$$Z_{nj} := [Y_{nj} - R_n(x_{nj})] \quad ; \tag{A1}$$

then $E(Z_{nj}) = 0$, $Var(Z_{nj}) = 1$, for $j = 0,1, \ldots, n$; and $Z_{no}, \ldots, Z_{nn}$ are independent. From (2.7) we have

$$U_{nr} = \sum_{j=0}^{n} \pi_{nj} r_{no}(x_{nj}) \phi_{nr}(x_{nj}) [Z_{nj} / \sqrt{\lambda_n \pi_{nj}}) + R_n(x_{nj}) - R_{no}(x_{nj})] ,$$

and from (2.1) and (2.4) we have

$$U_{nr} = (1/\sqrt{\lambda_n}) \sum_{j=0}^{n} \sqrt{\pi_{nj}} r_{no}(x_{nj}) \phi_{nr}(x_{nj}) Z_{nj} + \beta_{nr}, \ r = 0, \ldots, n .$$

Then from (2.10), (2.11), and (4.1),

$$V_{nr} = \sum_{j=0}^{n} \sqrt{\pi_{nj}} \phi_{nr}(x_{nj}) r_{no}(x_{nj}) Z_{nj} , \quad r = 0,1, \ldots, n . \tag{A2}$$

Let $f_{nj}$ be the characteristic function of $Z_{nj}$;

$$f_{nj}(t) := E[\exp(itZ_{nj})] , \ j = 0, 1, \ldots, n . \tag{A3}$$

Let $k_n$ be an increasing sequence of integers satisfying the hypotheses of Theorem 4.1, and let $\{a_{nj}, j = 0,1,\ldots, k_n ,$ $n = 0,1, \ldots \}$ be an array of real numbers such that $\sum_{j=0}^{k_n} a_{nj}^2 = 1$. Since $Z_{no}, \ldots, Z_{nn}$ are independent, it follows from (A2) that the characteristic function of $\sum_{r=0}^{k_n} a_{nr} V_{nr}$,

$$f_n^*(t) := E[\exp(it \sum_{r=0}^{k_n} a_{nr} V_{nr})] , \tag{A4}$$

is given by

$$f_n^*(t) = \prod_{j=0}^{n} f_{nj}([\sqrt{\pi_{nj}}\, r_{no}(x_{nj}) \sum_{r=0}^{k_n} a_{nr}\phi_{nr}(x_{nj})]t) .$$

For fixed $t$, set

$$t_{nj} := [\sqrt{\pi_{nj}}\, r_{no}(x_{nj}) \sum_{r=0}^{k_n} a_{nr}\phi_{nr}(x_{nj})]t ; \tag{A5}$$

then

$$f_n^*(t) = \prod_{j=0}^{n} f_{nj}(t_{nj}) . \tag{A6}$$

Since $E(Z_{nj}) = 0$ and $Var(Z_{nj}) = 1$, $j = 0,1,\ldots,n$,

$$f_{nj}(t) = 1 - t^2/2 + (\alpha_{nj}/6)|t|^3\, E|Z_{nj}|^3 ,$$

where $|\alpha_{nj}| <= 1$, and

$$f_{nj}(t_{nj}) = 1 + \theta_{nj} , \tag{A7}$$

where

$$\theta_{nj} := -t_{nj}^2/2 + (\alpha_{nj}/6)|t_{nj}|^3\, E|Z_{nj}|^3 . \tag{A8}$$

Since $|a_{nr}| <= 1$, $r = 0,1,\ldots,k_n$, by Hypothesis (4.3), we have

$$\max\{|t_{nj}| : j = 0,1,\ldots,n\} \longrightarrow 0 \text{ as } n \longrightarrow \infty , \tag{A9}$$

for fixed $t \in \mathbb{R}$. And according to (4.2), $E|Z_{nj}|^3 <= M$ for all $j$ and $n$, so that

$$\max\{|\theta_{nj}| : j = 0,1,\ldots,n\} \longrightarrow 0 \text{ as } n \longrightarrow \infty . \tag{A10}$$

From (2.4) and (A5) ,

$$\sum_{j=0}^{n} t_{nj}^2 = t^2 \sum_{j=0}^{n} \pi_{nj} r_{no}^2 (x_{nj}) \sum_{r=0}^{k_n} a_{nr}\phi_{nr}(x_{nj}) \sum_{s=0}^{k_n} a_{ns}\phi_{ns}(x_{nj})$$

$$= t^2 \sum_{r=0}^{k_n} a_{nr}{}^2,$$

so that

$$\sum_{j=0}^{n} t_{nj}^2 = t^2 \ . \qquad \text{(A11)}$$

Then

$$\sum_{j=0}^{n} |\theta_{nj}| <= t^2 / 2 + (M/6) \sum_{j=0}^{n} |t_{nj}|^3$$

$$<= t^2 / 2 + (Mt^2/6) \max\{|t_{nj}| : j = 0,1, \ldots , n\} \ ,$$

so that by (A9), for fixed $t$ ,

$$\sum_{j=0}^{n} |\theta_{nj}| \text{ is bounded.} \qquad \text{(A12)}$$

Again, from (A8) and (A11),

$$\left| \sum_{j=0}^{n} \theta_{nj} + t^2 / 2 \right| <= (Mt^2 / 6) \max\{|t_{nj}| : j = 0,1, \ldots , n| \ ,$$

so that by (A9)

$$\sum_{j=0}^{n} \theta_{nj} \text{ --> } -t^2/2 \text{ as } n \text{ --> } \infty \ .$$

It follows from (A6), (A7), (A10), (A12), and (A13) that

$$f_n^*(t) = \prod_{j=0}^{n} f_{nj}(t_{nj}) = \prod_{j=0}^{n} (1 + \theta_{nj}) \longrightarrow \exp(-t^2/2)$$

as $n \longrightarrow \infty$ (Chung, 1968, page 184), for each real $t$. So $\sum_{r=0}^{k_n} a_{nr} V_{nr}$ converges in law to the standard normal distribution, and the array $\mathcal{V}$ is jointly asymptotically normal. $\square$

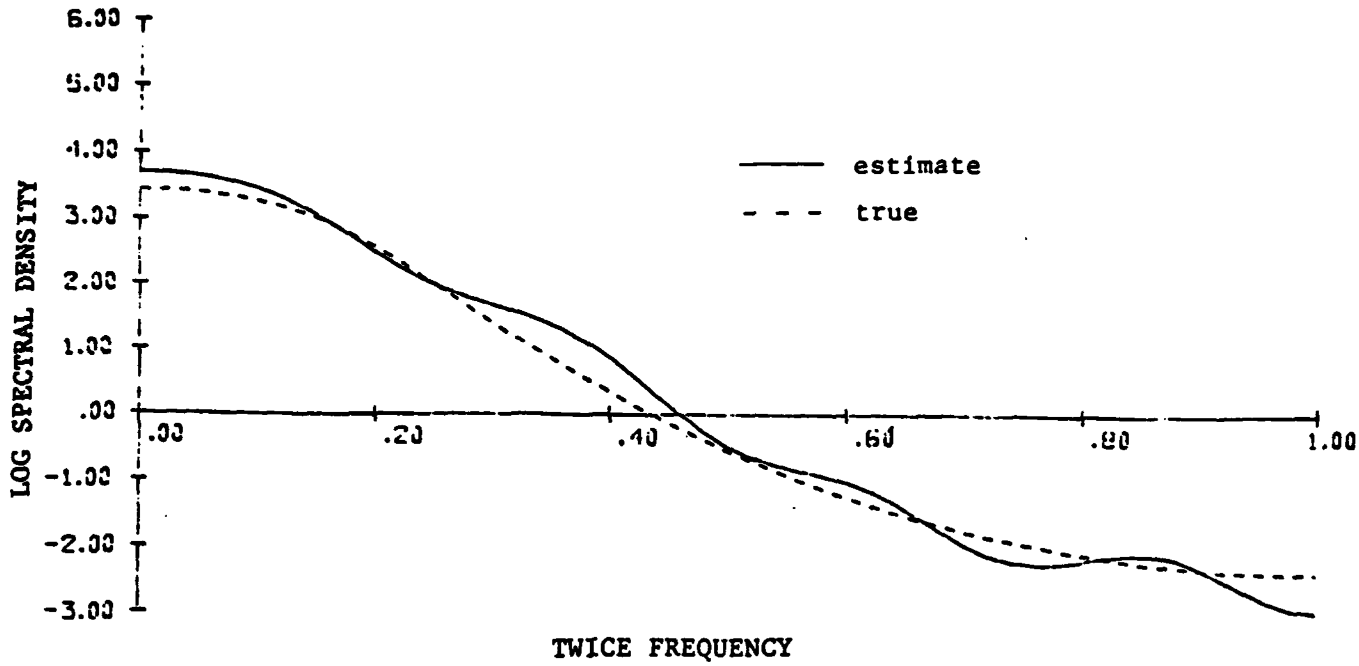



Figure 1. 256 points, first prior

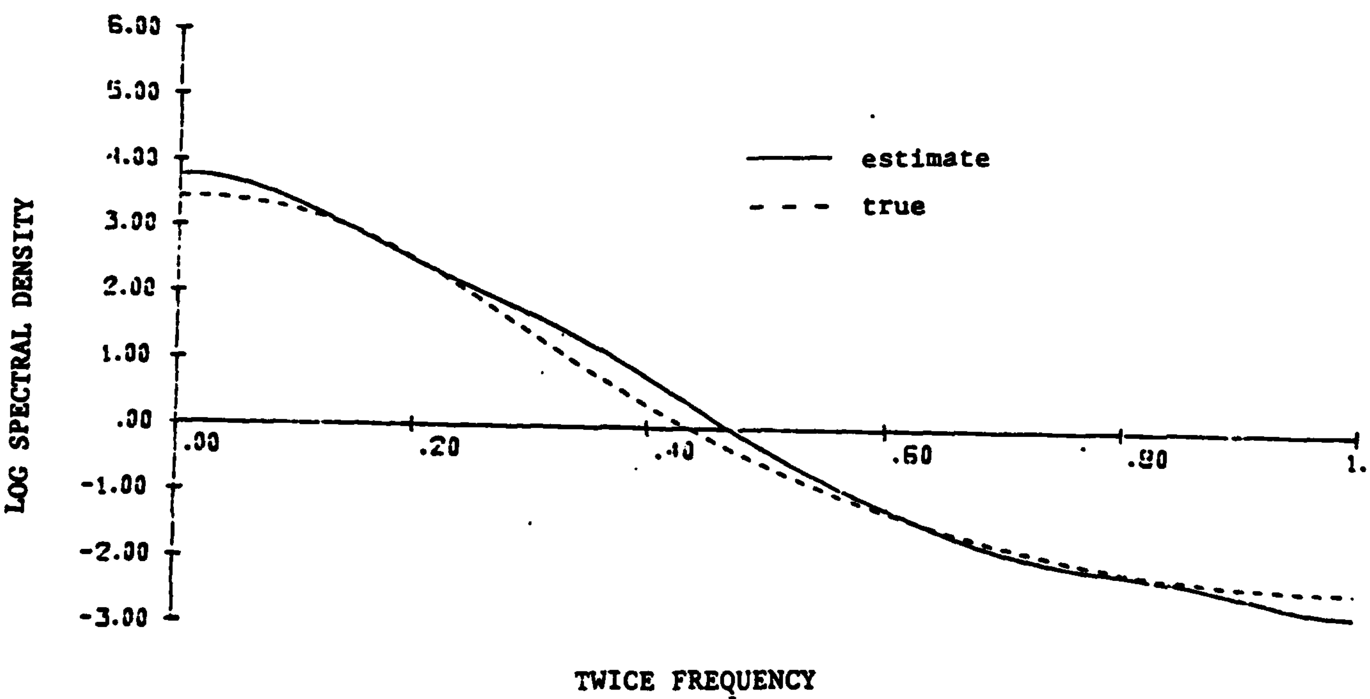



Figure 2. 256 points, second prior

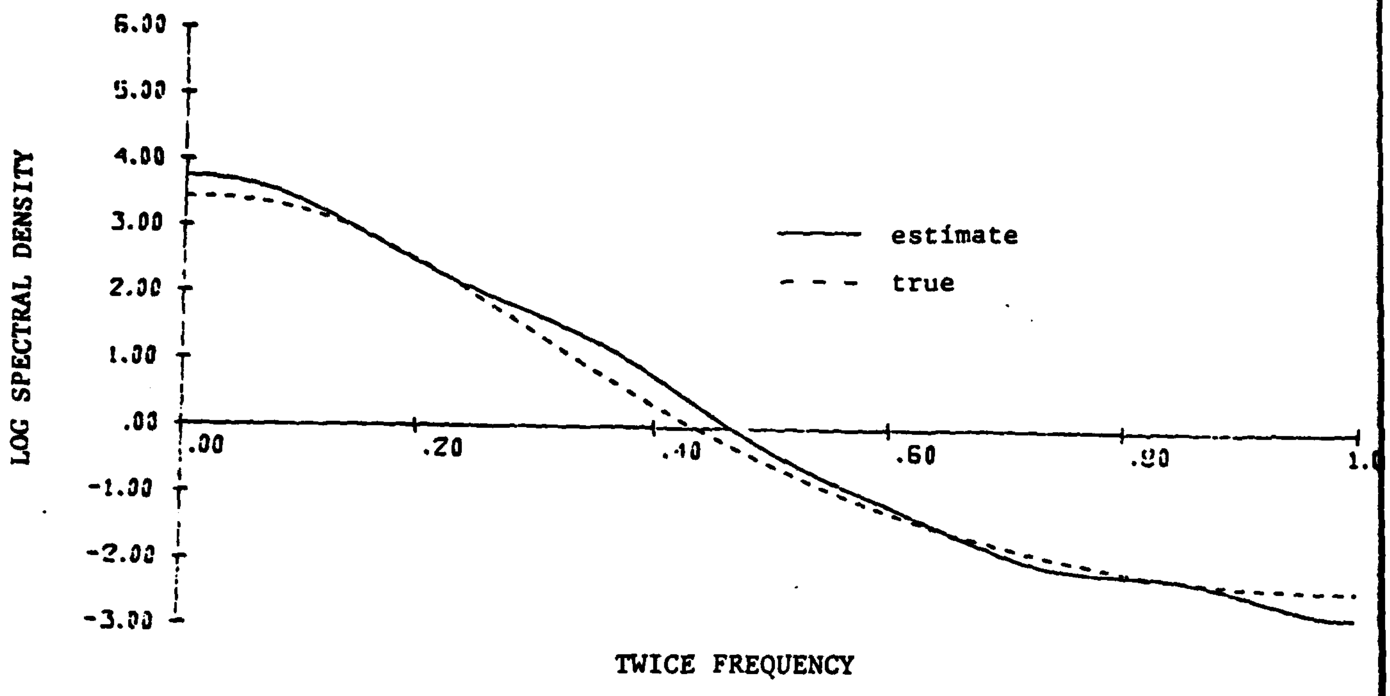



Figure 3. Third prior

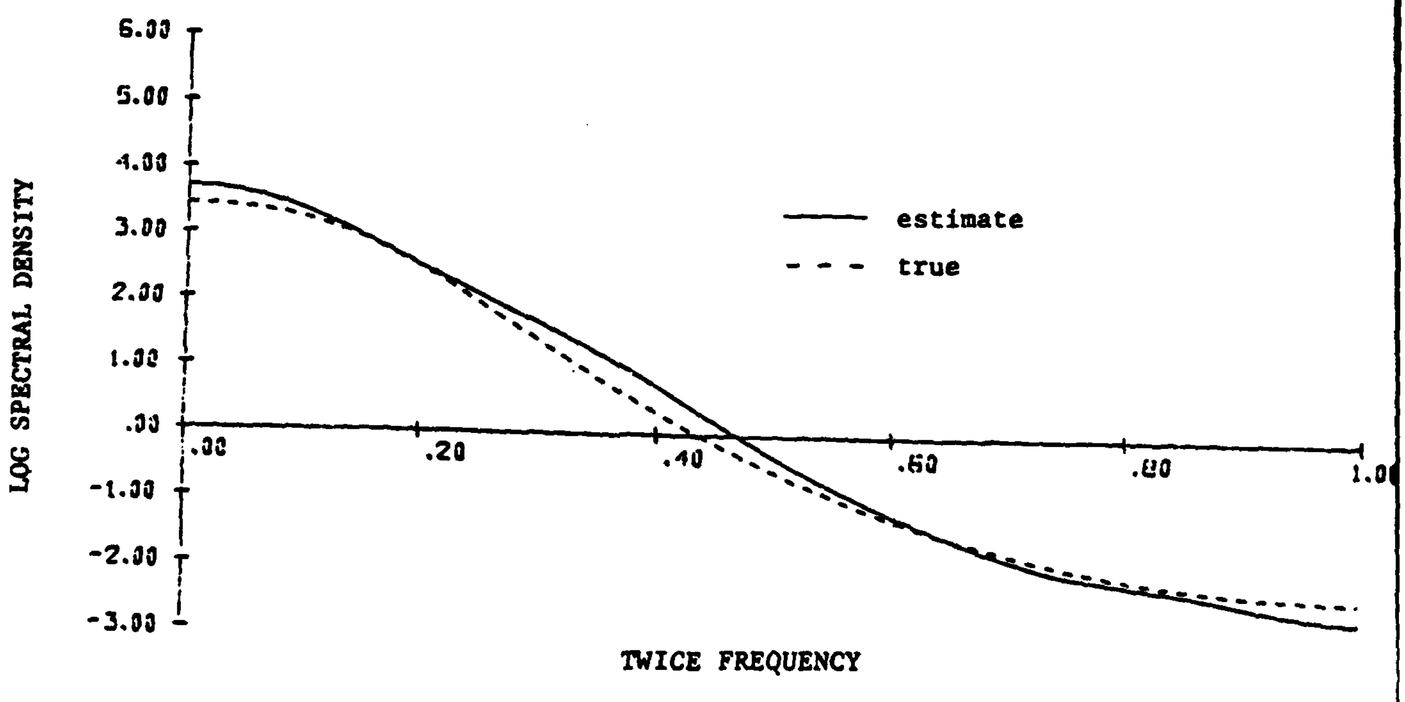



Figure 4. Four prior

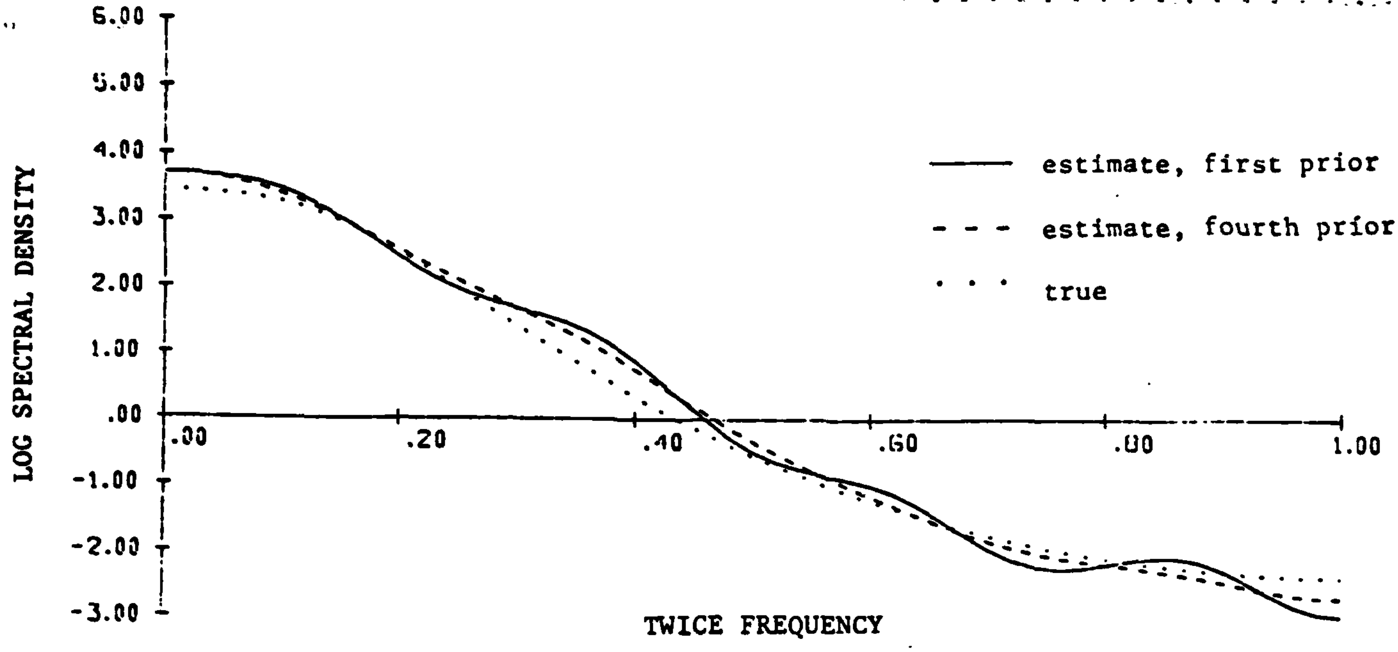



Figure 5. First and fourth priors.

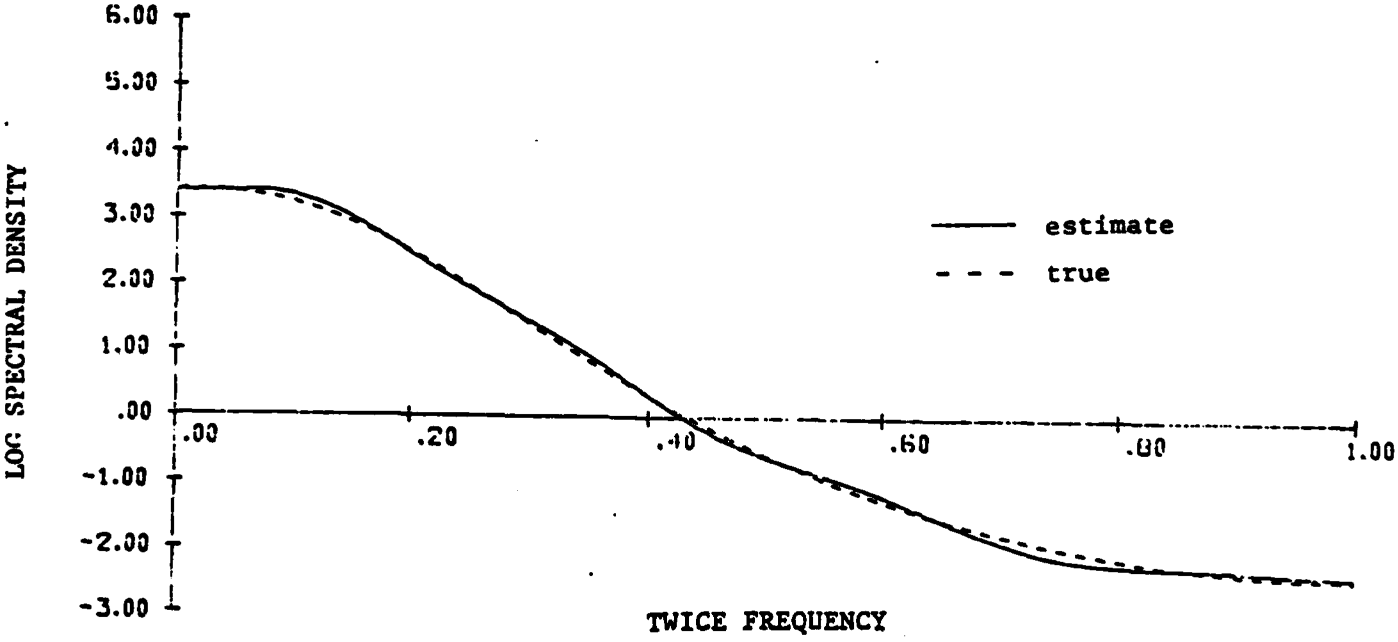



Figure 6. 1,024 points, first prior

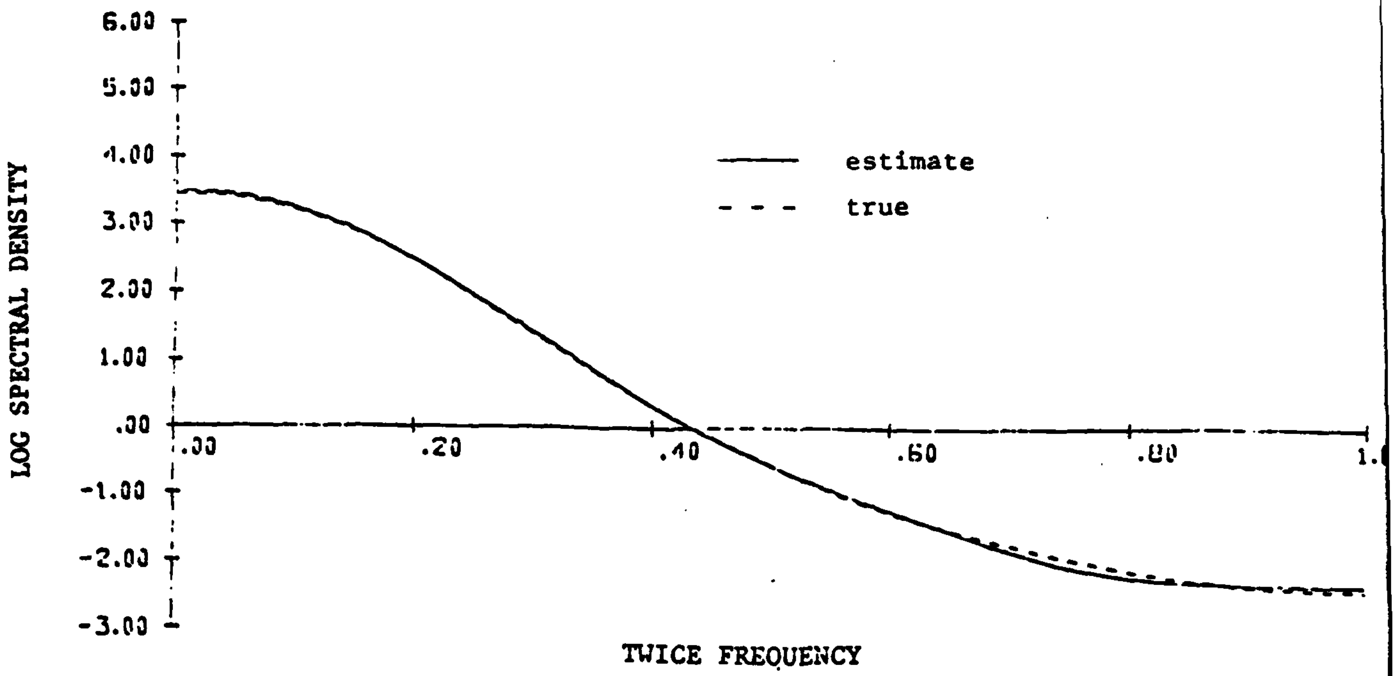



Figure 7. 1,024 points, fourth prior

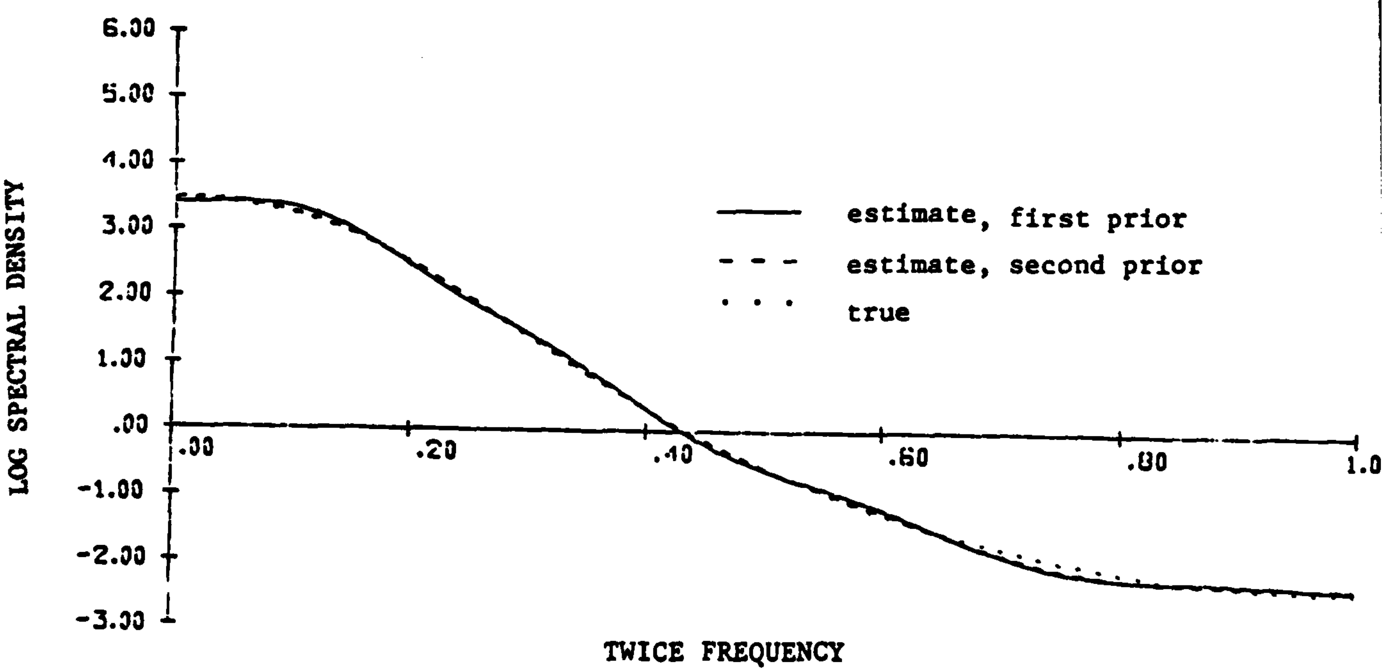



Figure 8. 1,024 points, first and fourth priors.

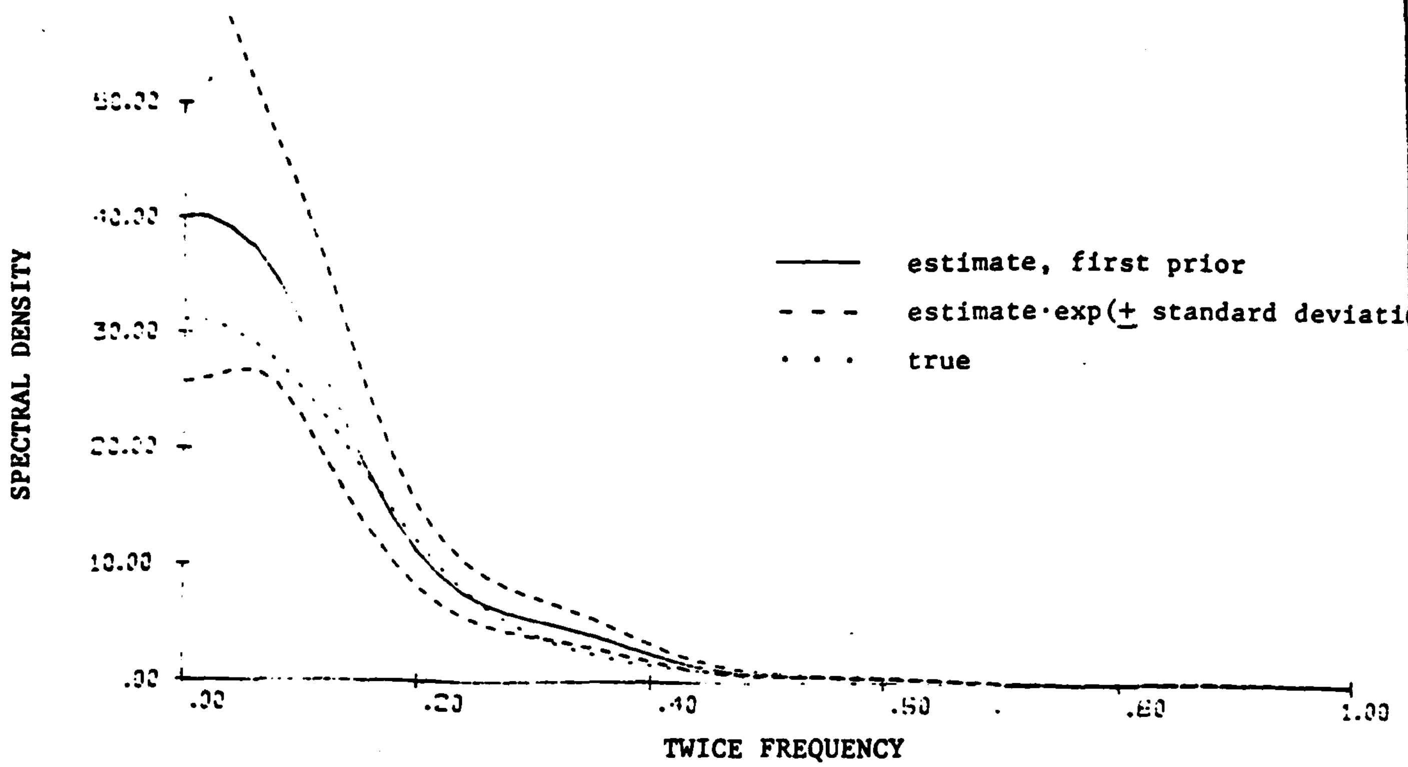



Figure 9. 256 points, first prior

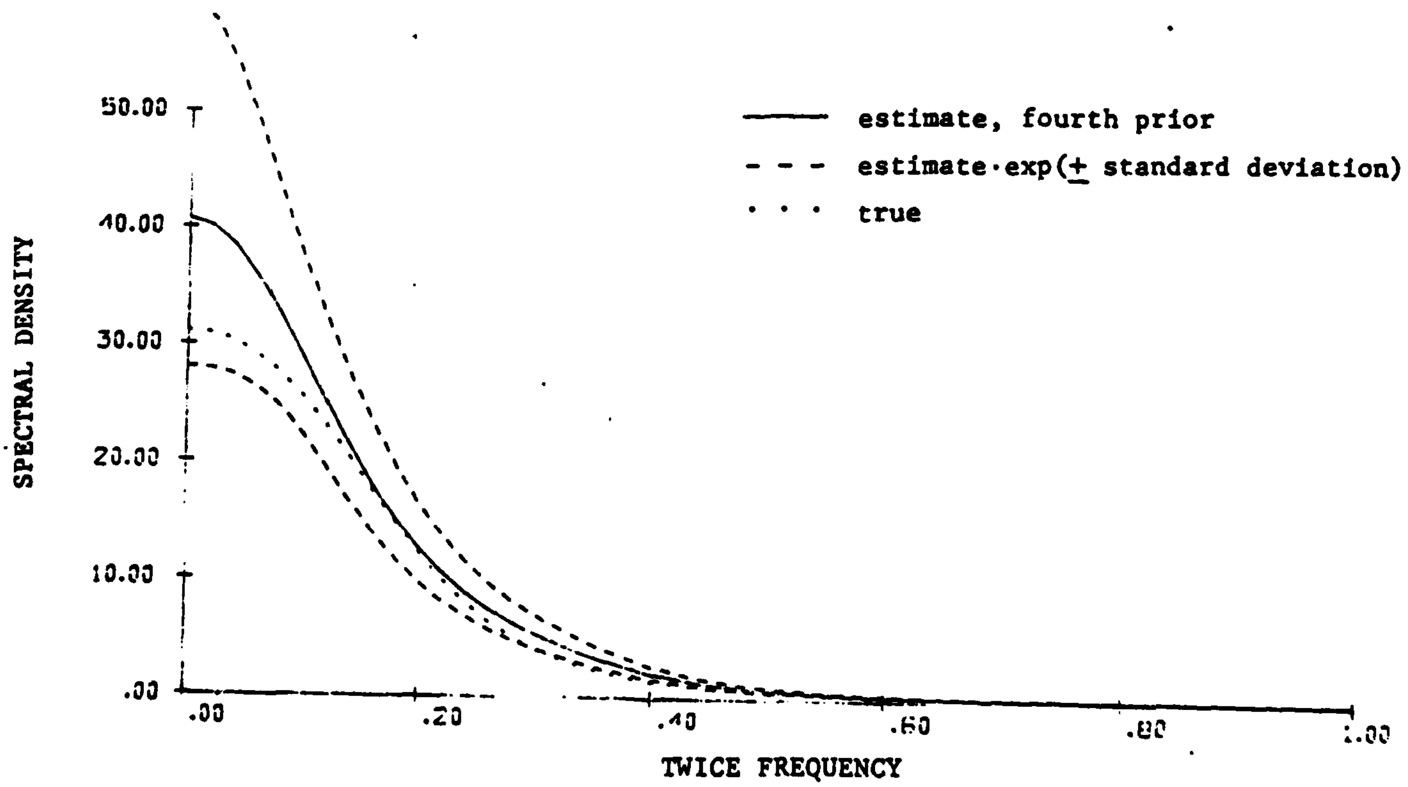



Figure 10. 256 points, fourth prior

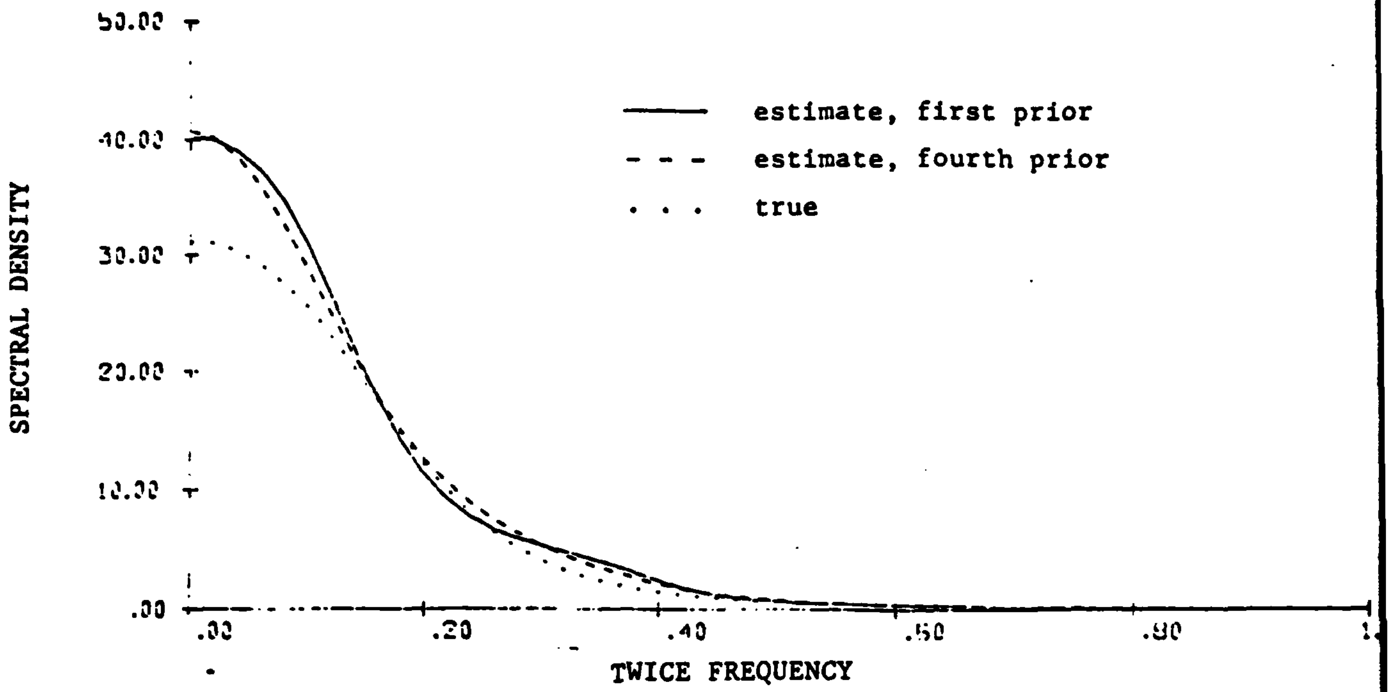



Figure 11. 256 points, first and fourth priors

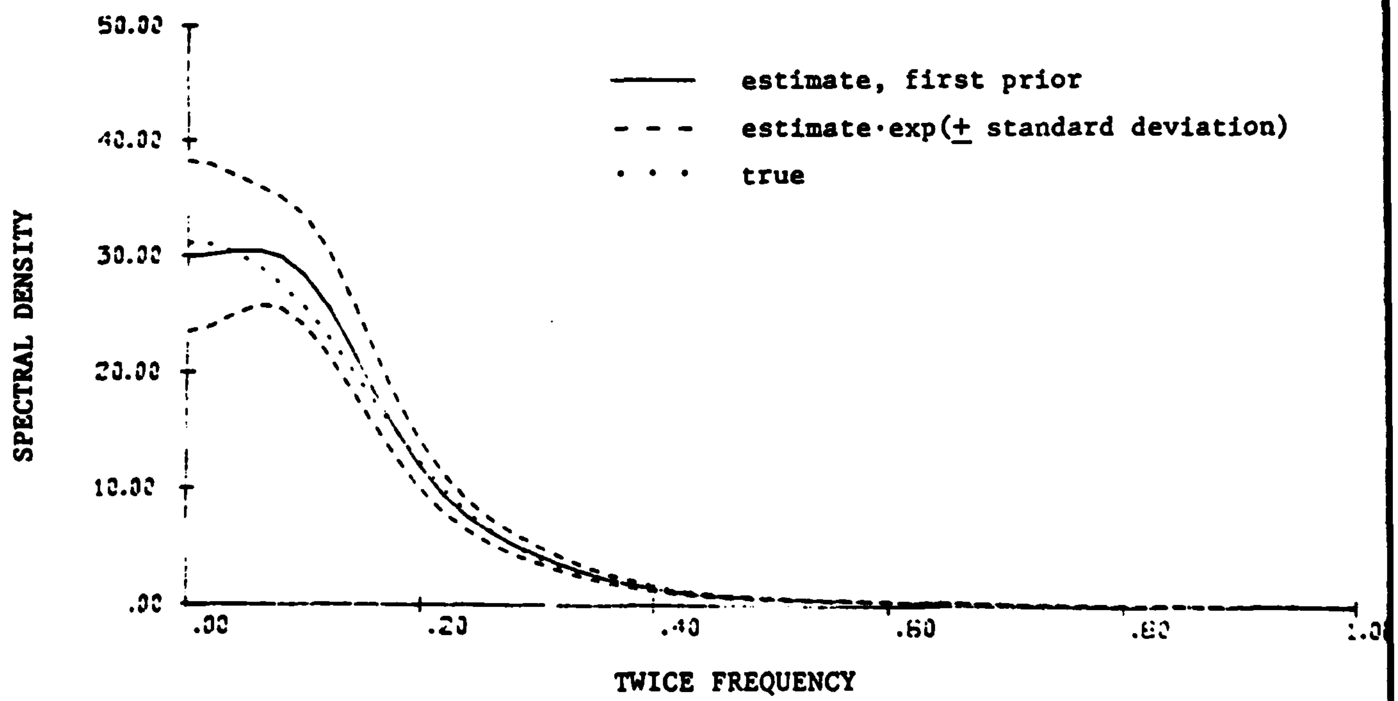



Figure 12. 1,024 points, first prior

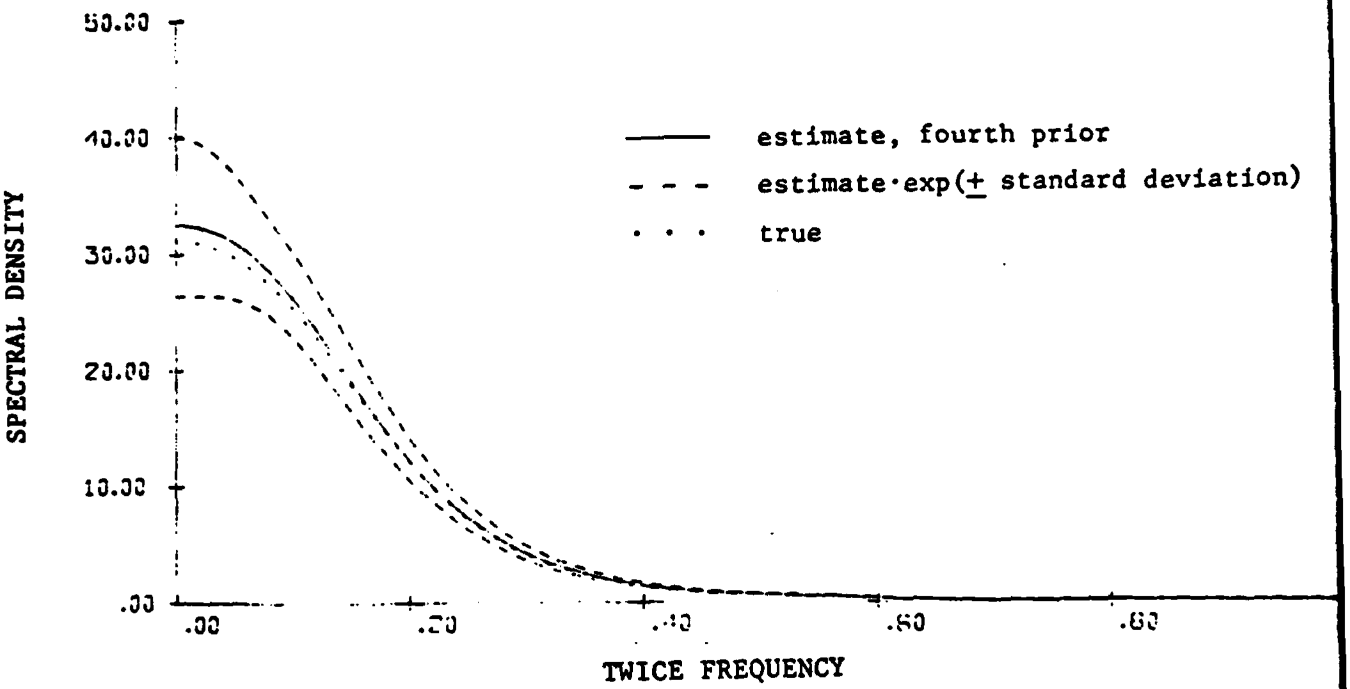



Figure 13. 1,024 points, fourth prior

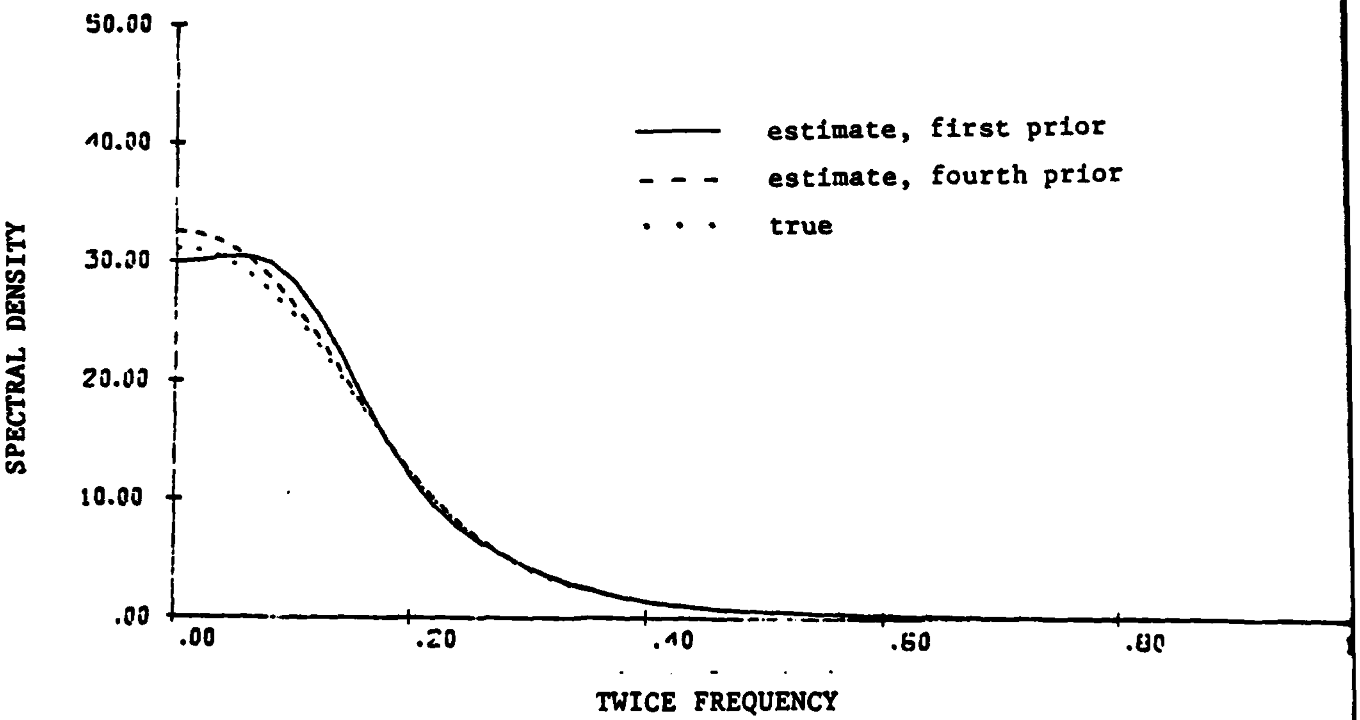



Figure 14. 1,024 points, first and fourth priors

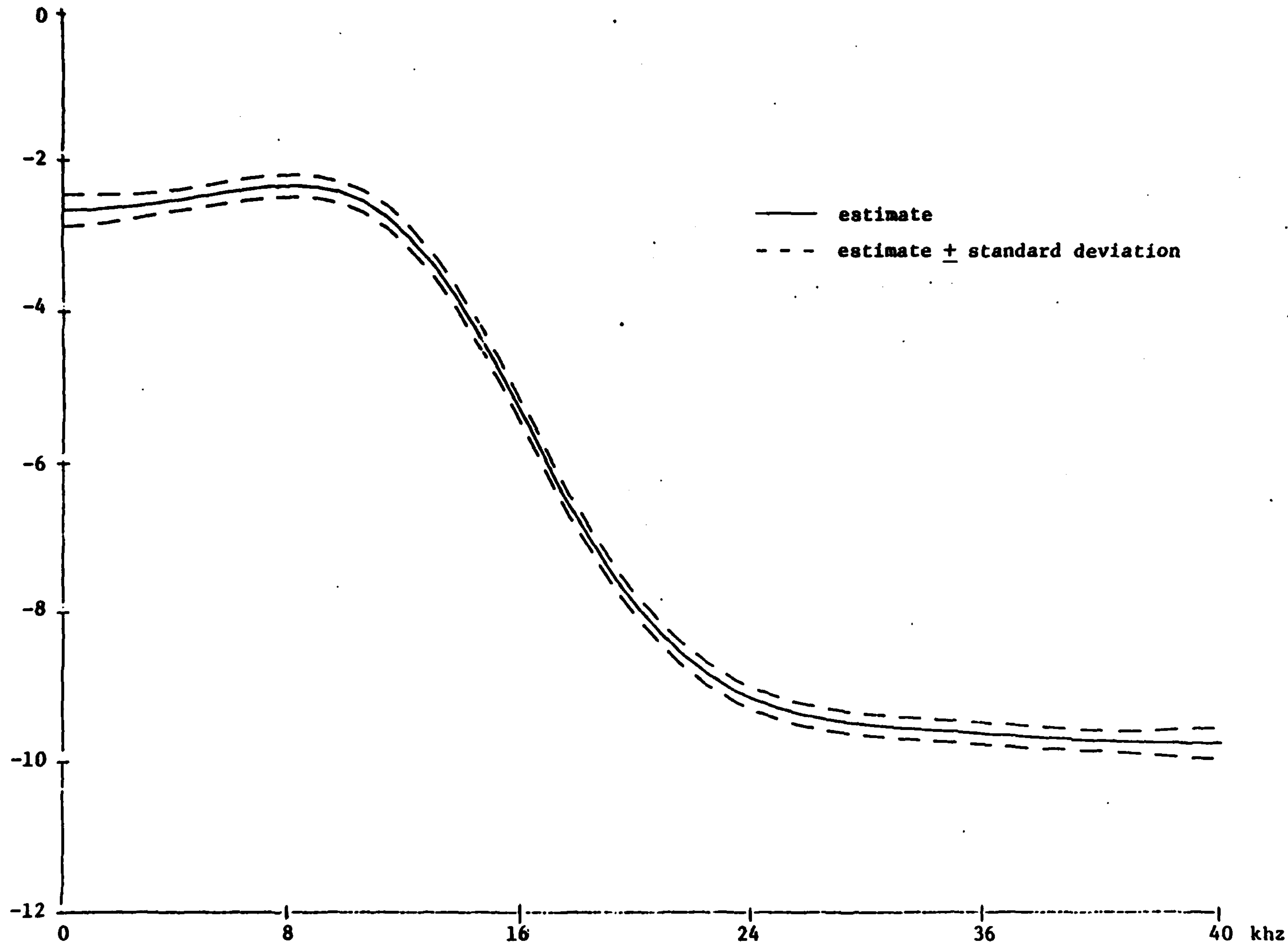



Figure 15. Log spectral density, ambient noise

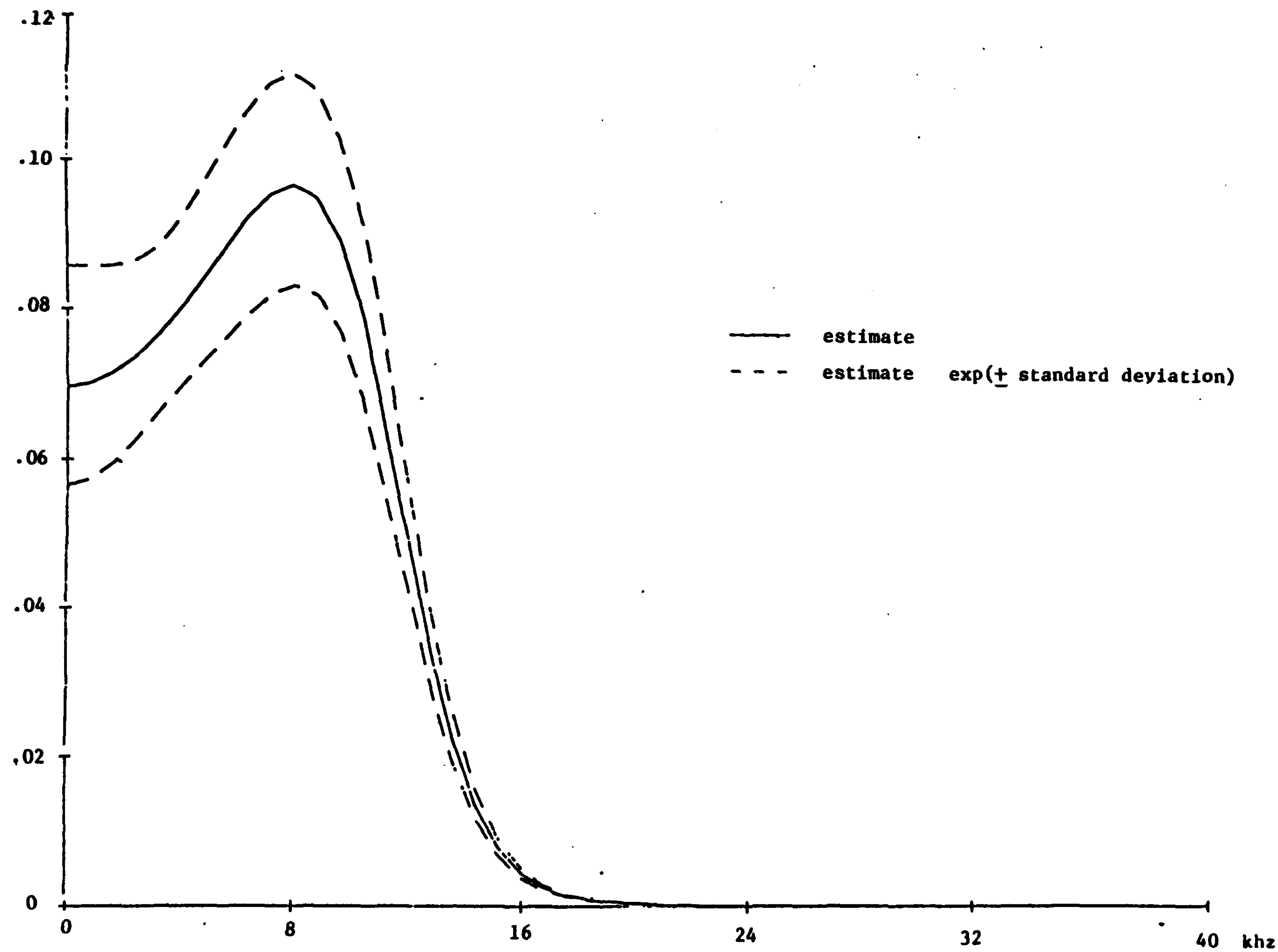



Figure 16. Spectral density, ambient noise

FILMED